

identify the candidate time. For instance, as described in more detail below with respect to step 206 of the method 200 (see Figure 2) the score associated with a candidate time can be adjusted based on the presence or absence of one or more cues within a specified time window that includes the candidate time or to which the candidate time is sufficiently proximate (i.e., is less than a specified short amount of time, such as several seconds, before or after the time window). The score for a candidate time can also be adjusted, for example, based on an evaluation of the relationship between the candidate time and one or more other candidate times. In particular, as described in more detail below with respect to step 207 of the method 200 (see Figure 2), this latter type of adjustment can make use of one or more probability models that describe expected relationship(s) between a candidate time and the one or more other candidate times.

Please replace the paragraph beginning at page 11, line 16, with the following rewritten paragraph:

In step 205 of the method 200, one or more of the cues identified in step 204 are analyzed to identify candidate times within the audiovisual content at which a commercial beginning or a commercial ending may occur. For example, an audio pause often accompanies either the beginning or the end of a commercial, so the presence of an audio pause in the audio content can be identified as a factor that militates toward establishing a candidate time at some time

during or proximate to the audio pause. Similarly, a sequence of black frames often accompanies either the beginning or the end of a commercial, so the presence of a sequence of black frames in the visual content can be identified as a factor that militates toward establishing a candidate time at some time during or proximate to the sequence of black frames. A scene cut or fade also typically accompanies the beginning or the end of a commercial, so the presence of a scene break or fade in the visual content can be identified as a factor that militates toward establishing a candidate time at some time during or proximate to the scene break or fade. The beginning and end of a commercial break are often accompanied by a noticeable increase and decrease in volume, respectively, so that a significant change in average volume (measured over a specified window of time) can be identified as a factor that militates toward establishing a candidate time at some time proximate to times at which the volume is seen to change significantly. Commercials often include relatively more musical content than the rest of a set of audiovisual content, so the occurrence of a time window of specified duration (e.g., the expected duration of a typical commercial break, such as 60 seconds, or the expected duration of a typical commercial, such as 15 or 30 seconds) having relatively high musical content (e.g., relatively high density of musical content relative to the density of musical content in other, proximate time windows) can be

identified as a factor that militates toward establishing candidate times at the beginning and end of such a time window. The beginning or end of a commercial is often accompanied by a change in speaker identity, so the occurrence of a change in speaker identity can be identified as a factor that militates toward establishing a candidate time at, or proximate to, the time at which such a change in speaker identity occurs. A commercial break often includes a relatively high density of scene breaks and/or fades (since a scene break or fade typically occurs at the beginning and end of a commercial break, as well as at the transition between commercials within a commercial break, and since commercials often include a relatively large number of scene breaks and/or fades per unit time within the commercial), so the occurrence of a time window of a specified duration (e.g., 60 seconds) during which the density of scene breaks and/or scene fades is relatively high (i.e., exceeds a specified threshold), or a significant change in density of scene breaks and/or scene fades over one window of time with respect to a proximate window of time, can be identified as a factor that militates toward establishing candidate times at the beginning and end of such a time window. A network icon is sometimes present during the noncommercial parts of a television broadcast; therefore, if a network icon is determined to be present in a set of audiovisual content, the disappearance of the network icon typically accompanies the beginning of a

commercial break and the appearance of the network icon typically accompanies the end of a commercial break, so the appearance or disappearance of a network icon can be identified as a factor that militates toward establishing a candidate time at, or proximate to, a time at which the network icon appears or disappears. Since the average motion level in the visual content of a commercial is often significantly different from the average motion level of other visual content in a set of audiovisual content, significant change in the amount of motion in the visual content of a time window (e.g., about 60 seconds) relative to the amount of motion in the visual content in a proximate time window can be identified as a factor that militates toward establishing candidate times at, or proximate to, the beginning and end of such a time window. The appearance of text (other than closed-captioning) in a set of audiovisual content often accompanies the beginning of a commercial break and the disappearance of text often accompanies the end of a commercial break, so the appearance or disappearance in a set of audiovisual content of text other than closed-captioning can be identified as a factor that militates toward establishing a candidate time at, or proximate to, a time at which text appears or disappears. If closed-captioning data is present in the data representing the audiovisual content, a closed-captioning scrolling format change often occurs at the beginning or the end of a commercial break, so the occurrence of a

closed-captioning scrolling format change can be identified as a factor that militates toward establishing a candidate time at, or proximate to, the time at which such a format change occurs. If closed-captioning data is present in the data representing the audiovisual content, the disappearance of closed-captioning often accompanies the beginning of a commercial break and the appearance of closed-captioning often accompanies the end of a commercial break, so the appearance or disappearance of closed-captioning can be identified as a factor that militates toward establishing a candidate time at, or proximate to, a time at which closed-captioning appears or disappears.

Please replace the paragraph beginning at page 14, line 9, with the following rewritten paragraph:

As indicated above, it is an advantageous aspect of the invention that the invention enables use of a combination of the cues to effect commercial detection. In particular, the invention can enable the use of cues and combinations of cues that have not previously been used for commercial detection. For example, the invention can advantageously enable any one of detection of the absence of a network icon, an analysis of musical content present in a set of audiovisual content, the density of scene breaks and/or fades, or analysis of the identity of speakers of spoken content to be used alone as a commercial detection cue. These cues can also be used in any combination with each other or any other cue. In particular, it is anticipated

that one or more of these cues can advantageously be used in combination with one or more of the following cues: 1) the occurrence of an audio pause, 2) the occurrence of a sequence of black frames, 3) a scene cut or fade, 4) the occurrence of specified closed-captioning formatting signals, and 5) the appearance or disappearance of closed-captioning.

Please replace the paragraph beginning at page 14, line 28, with the following rewritten paragraph:

Step 205 outputs a list of candidate times at which commercials may be beginning or ending, together with a score or probability associated with each candidate time. In one implementation of the invention, each candidate time is assigned the same initial score. Alternatively, the scores assigned to candidate times can vary. For example, the score for a candidate time can depend on which cue(s) were used to identify the candidate time. The beginning or end of a commercial can be deduced from the presence of some cues with a greater degree of confidence than that associated with the presence of other cues. To the extent that a candidate time is identified based on a cue with which a relatively high degree of predictive confidence is associated, the score assigned to that candidate time can be relatively higher than would be the case if a relatively low degree of predictive confidence was associated with the cue. Additionally, the score for each candidate time can be dependent on how strongly the cue is present in the

audiovisual content, as determined in accordance with a criterion or criteria appropriate for that cue: the more strongly a cue is present, the higher the score. For example, when one of the cues used to establish a candidate time is an audio pause, the score established for the candidate time can be dependent on the duration of the audio pause and/or the degree of silence during the audio pause (e.g., the score for the candidate time is made relatively greater the longer the audio pause or the less sound that is present during the audio pause). Or, for example, when one of the cues used to establish a candidate time is a sequence of black frames, the score established for the candidate time can be dependent on the duration of the sequence of black frames and/or the completeness of the blackness of the frames (e.g., the score for the candidate time is made relatively greater the longer or blacker the sequence of black frames). Or, for example, when one of the cues used to establish a candidate time is a scene cut, the score established for the candidate time can be dependent on the number of pixels that changed by more than a threshold amount from one frame to another (e.g., the score for the candidate time is made relatively greater as more pixels changed between scenes) and/or dependent on the total change of all the pixels from one frame to another (where the "change" for each pixel is the change in the color or other components of a pixel). Or, for example, when one of the cues used to establish a candidate time is a significant

average audio volume change, the score established for the candidate time can be dependent on degree of the volume change (e.g., the score for the candidate time is made relatively greater as degree of the volume change increases). Those skilled in the art can readily appreciate how the score for a candidate time can be adjusted based on aspects of other cues present in the audiovisual content proximate to the candidate time. Additionally, the score for a candidate time can be dependent on the confidence level associated with identification of the cue in the audiovisual content: the greater the confidence level, the higher the score. (This confidence level is different from the confidence level associated with the predictive capability of the cue, discussed above.) For example, sound represented in audio data may be sound in the audio content or noise. The score for a candidate time identified at least in part based on the presence of an audio pause can be increased or decreased in accordance with extent to which the degree of noise present in the audio data increases or decreases the confidence with which an audio pause can be detected.

Please replace the paragraph beginning at page 16, line 22, with the following rewritten paragraph:

In step 206 of the method 200, the scores associated with each candidate time can be adjusted based on the presence or absence of one or more cues within some time window proximate to the candidate time. The cue(s) used to



adjust the score of a candidate time in step 206 are different from the cue(s) used to establish the candidate time and an initial associated score in step 205. The duration of the time window and location of the time window with respect to the candidate time is dependent on the type of cue. For instance, the score for a candidate time is increased (i.e., the likelihood that the candidate time correctly indicates the beginning or ending of a commercial is deemed to increase) in each of the following cases: 1) the candidate time is coincident with the time at which an audio pause (which is a window of audio silence or near silence) occurs, 2) the candidate time is within or sufficiently proximate to a time window in which the closed-captioning scrolling format is different from that which is typical for audiovisual content of this type, 3) the candidate time is within or sufficiently proximate to a time window during which closed-captioning is absent (for audiovisual content that is known to be closed-captioned), 4) the candidate time is within or sufficiently proximate to a time window of at least a specified duration (e.g., 60 seconds) and including high musical content, 5) the candidate time is within or sufficiently proximate to a time window during which the density of scene breaks and/or scene fades exceeds a specified threshold, 6) the candidate time is sufficiently proximate to a time window of at least a specified duration (e.g., 0.5 seconds) and in which the average motion in the

visual content, measured in a specified manner, is less than a specified threshold, 7) the candidate time is within a time window during which a network icon (which has been found to be persistent through a majority of the visual content) is not present at a specified location within the visual content (e.g., a region, such as a corner, near the edge of the visual content), 8) the candidate time is very near (e.g., within about 2 seconds) a time at which the time-averaged audio volume (averaged over a time window of about 10 seconds) has changed by a magnitude of greater than a specified threshold, 9) the candidate time is sufficiently proximate to (within about 1 second) a time when text is present in the visual content, 10) the candidate time is within a specified duration of time (e.g., a few seconds) after the presence in the closed-captioning stream of certain keywords or phrases such as "commercial", "break", "coming up" or "after", or within a specified duration of time (e.g., a few seconds) prior to the presence in the closed-captioning stream of certain keywords or phrases such as "welcome", "hello" or "we're back, 11) the candidate time is within a specified duration of time (e.g., 2 seconds) from a time at which the speaker identity has changed, and 12) the candidate time is within a specified duration of time (e.g., one to several seconds) from a time window of greater than a specified duration (e.g., 1 minute) that does not include speech from a speaker whose speech has been determined to be present in the audiovisual content with

greater than a specified frequency. The amount by which a score is adjusted can be dependent on the same types of analyses done to establish an initial score for a candidate time, as described above with respect to step 205.

(However, the particular analyses done in step 206 need not, but can, be the same as those done in step 205.) In particular, the amount of the adjustment to a score for a candidate time can be dependent on how strongly the cue is present in the audiovisual content, as determined in accordance with a criterion or criteria appropriate for that cue: in general, the more strongly a cue is present, the greater the adjustment to the score. Additionally, the amount of the adjustment to a score for a candidate time can be dependent on how high or low the score is prior to the adjustment. For example, a cue that strongly indicates the presence of a commercial beginning or ending may cause a larger adjustment in a relatively low score than in a relatively high score. The particular quantities, keywords, and other algorithm parameters given above are illustrative; they may be changed, within appropriate constraints, as can be appreciated by those skilled in the art, without adversely affecting the operation of the invention.

Please replace the paragraph beginning at page 22, line 9, with the following rewritten paragraph:

Step 208 begins by selecting the candidate time with the highest score to be a commercial start or end time (whether that time is a start time or end time is unknown at

this point). A commercial break is then constructed based on the selected candidate time by successively evaluating candidate times in order of decreasing score and adding candidate times to the commercial break that satisfy each of the following criteria: 1) the additional candidate time is well-spaced in time, in accordance with the function  $S(t)$ , from each candidate time that has already been included in the commercial break, 2) the additional candidate time does not create a commercial break which is too long, in accordance with the function  $L(t)$ , and 3) the additional candidate time is not too close to other existing commercial breaks, in accordance with the function  $W(t)$ , that have already been identified by the step 208. Stated another way, candidate times continue to be added to a commercial break in order of score as long as there are any candidate times for which all of the following are true: 1) the value of  $S(t)$ , where " $t$ " is the time separation between the candidate time being evaluated and a candidate time already included in the commercial break, is above a specified threshold value for each candidate time already included in the commercial break, 2) the value of  $L(t)$ , where " $t$ " is the duration of the commercial break if the candidate time is added to the commercial break, is above a specified threshold value, and 3) the value of  $W(t)$ , where " $t$ " is the time separation between the candidate time and an existing commercial break, is above a specified threshold value for each existing commercial break.

Please replace the paragraph beginning at page 26, line 25, with the following rewritten paragraph:

The invention can be used for a wide variety of applications, as can be appreciated by those skilled in the art in view of the description herein. In general, the invention can be used with any broadcast or other data transmission over a network (e.g., conventional network television broadcasts, cable television broadcasts, broadcasts or transmissions over a computer network such as the Internet - and, in particular, the World Wide Web portion of the Internet). Additionally, the invention can be used generally to detect commercials in audiovisual content represented by any type of data, which data can be stored on a data storage medium or media, or provided to a system or method according to the invention in real time. Further, the invention can be implemented in a wide variety of apparatus, as can also be appreciated by those skilled in the art in view of the description herein, such as, for example, television set-top boxes, digital VCRs, computers (including desktop, portable or handheld computers) or any of a variety of other computational devices (including many which are now being, or will in the future be, developed).